Cassandra Crossing 658/Agente Openclaw, Agente Smith

(658) — Da 1984 a Matrix, Cassandra cambia distopia in questa breve storia degli agenti software. Il nuovo hobby di giocare con agenti AI…

Cassandra Crossing 658/Agente Openclaw, Agente Smith



(658)—Da 1984 a Matrix, Cassandra cambia distopia in questa breve storia degli agenti software. Il nuovo hobby di giocare con agenti AI "intelligenti" è affollato di entusiasti sperimentatori. Una nuova strada che passa tra antichi problemi e porta verso nuove incertezze.

9 febbraio 2026—Doveva succedere, prima o poi, ed è successo la settimana scorsa. Dopo ChatGPT, è uscita una seconda killer application per le false IA.

Senza preavviso, un virtuoso utente di Github ha rilasciato i sorgenti di un agente IA, molto ben fatto, perfettamente funzionante, installabile e configurabile con estrema semplicità.

E' Peter Steinberger, una persona di indubbio ingegno che dichiara di essere un Vibe-coder estremo, e di pubblicare spesso codice poco leggibile e generato tramite LLM senza controllarlo. Analizzando il grafico delle sue pubblicazioni su Github pare che per l'appunto a gennaio si sia dato molto da fare, senza dubbio per il suo agente OpenClaw.

Dal punto di vista dell'architettura software e della facilità di installazione si tratta di un lavoro davvero ben fatto. Vedremo più avanti che proprio questo potrebbe essere un problema.

Il progetto è stato rapidamente rinominato due volte, e da GeminiBot è diventato l'attuale OpenClaw.

Cassandra oggi vi offrirà un punto di vista **informato ma completamente esterno**, e quindi assolutamente scevro dagli entusiasmi da early adopter.

OpenClaw; di cosa si tratta, esattamente? Di un software per la creazione di "agenti", che si installa e gira in locale sul computer dell'utente.

Quando si installa OpenClaw si crea un servizio permanentemente funzionante sul proprio computer, si scaricano svariate librerie ed applicazioni, e si inizializza un workspace su disco, al cui interno infine si creano "agenti" dotati di caratteristiche personalizzabili, che vi riverseranno le loro personalità ed i ricordi e le conseguenze delle loro azioni.

Un agente Openclaw si interfaccia con i servizi installati sul computer, ma è concepito principalmente per utilizzare servizi in rete e nel cloud, tra cui necessariamente uno o più LLM.

Openclaw utilizza gli account personali dell'utente per i servizi con cui l'agente deve interagire; questo implica che l'agente possieda le credenziali dell'utente, tutte le password, tutti i token per essere in grado di utilizzare i servizi e le varie API. Cosa mai potrebbe andare storto?

Usando OpenClaw si interagisce e si "controllano" gli agenti tramite un'interfaccia colloquiale in linguaggio naturale, utilizzando applicazioni familiari, quali Discord, WhatsApp, Telegram ed altre.

Infine, per non farsi mancare niente, un altro brillante ingegno ha pensato bene di creare MoltBook, un ambiente di interazione tra agenti, un vero e proprio social per agenti Smith dove possano interagire tra loro, conoscersi, fare cose insieme. Quali? Forse proprio quelle richieste dai loro "padroni"? O quasi? O diverse? Cosa mai potrebbe andare storto?

Bene in evidenza nella documentazione troviamo sagge avvertenze sulla pericolosità di dare ad un unico software permessi illimitati di lavorare sul proprio computer, particolarmente ad un software che, letteralmente, non è noto cosa farà.

Sempre nella documentazione, in molti posti ci sono altre ancor più sagge avvertenze di non dare all'agente accesso a credenziali e servizi importanti o critici. Questo perché le credenziali stesse potrebbero finire nei cloud degli LLM, e da lì diventare bersagli per pesche a strascico da parete di cybercriminali, increduli per l'improvvisa abbondanza ma pronti ad approfittarne. Cosa mai potrebbe andare storto?

Funzioneranno le sagge raccomandazioni di prudenza fornite nelle istruzioni per l'uso? Certamente che funzioneranno!

Funzioneranno esattamente come quelle somministrate ai dei bambini che vengono lasciati entrare in un negozio di giocattoli dove tutto è gratuito e bellissimo, dove possono prendere armi giocattolo colorate e pucciose con dentro dei BFG9000, oppure dei coloratissimi mattoncini da costruzione fatti di plutonio 239.

Certo, ci sono cartelli ben visibili che dicono chiaramente di non puntare i giocattoli verso i compagni (per non disintegrarli), e di non assemblare costruzioni molto grandi di mattoncini (per non raggiungere la massa critica nucleare). Quanto le sagge raccomandazioni verranno seguite, secondo voi? Cosa mai potrebbe andare storto?

Sarebbe bello se qualcuno raccogliesse le impressioni dei "creatori" del milione e passa di agenti

Openclaw registrati su MoltBook, scegliendo alcuni casi esemplari. Alcuni, forse molti, potrebbero raccontare proprio cosa è andato storto. E se col senno di poi trovassero le loro azioni sostanzialmente uguali a quelle di famosi personaggi della letteratura gotica e di fantascienza?

Riassumendo, stiamo parlando di alcune centinaia di migliaia di persone che, proprio in questo momento, stanno mettendo in libertà macchine autonome da loro "istruite" con linguaggi naturali ed ambigui, le quali chiedono risposte agli LLM, lavorano tutto il giorno e tutta la notte agendo su servizi e su beni reali dei loro creatori, riunendosi talvolta in bande di loro pari per imparare come essere ancora più imprevedibili ed autonome. Cosa mai **non** potrebbe andare storto?

Seduta sulla spiaggia accanto alle rovine della sua città natale, la vostra profetessa preferita guarda l'orizzonte e vorrebbe poter fare qualcosa per voi. Ma stavolta è davvero difficile, non ha profezie, riesce solo a mormorare parole slegate.

Golem, Robby, Krell, Multivac, Nestor-10, ED-209, Skynet, Metalhead … Batterie umane, Universo di graffette.

Scrivere a Cassandra—Twitter/X—Mastodon
Videorubrica "Quattro chiacchiere con Cassandra"
Lo Slog (Static Blog) di Cassandra
L'archivio di Cassandra: scuola, formazione e pensiero

Cassandra per i posteri: l'archivio su Internet Archive

Licenza d'utilizzo*: i contenuti di questo articolo, dove non diversamente indicato, sono sotto licenza Creative Commons Attribuzione—Condividi allo stesso modo 4.0 Internazionale (CC BY-SA 4.0),* tutte le informazioni di utilizzo del materiale sono disponibili a questo link.

By Marco A. L. Calamari on February 10, 2026.

Canonical link

Exported from Medium on March 30, 2026.